Probing the Molecular Basis of Protein Function through Chemistry

Roelfes, Gerard; Mootz, Henning D.

Published in:
EPRINTS-BOOK-TITLE

Publication date:
2006

DOI: 10.1002/cbic.200500547

Probing the Molecular Basis of Protein Function through Chemistry

Gerard Roelfes*[a] and Henning D. Mootz*[b]

In the post-genomic era, with the sequence information of entire genomes available, we are still far from understanding the complexity of the proteome. Proteins orchestrate the complex and delicate biological processes in living cells, yet fundamental questions about protein structure and function remain open. For example, how are protein structure and function encoded by the amino acid sequence and how do the myriad of post-translational modifications influence their properties? To discuss how chemistry can help in expanding our knowledge of protein function, a group of 80 international scientists gathered in the small town of St. Feliu de Guixols, on the Costa Brava in Spain. The occasion was the ESF/EMBO conference on "Probing the Molecular Basis of Protein Function through Chemistry", organized by Martin Engelhard (MPI Dortmund) and Don Hilvert (ETH Zürich).

Technological advances have significantly improved our ability to apply synthetic chemistry to large biomolecules, such as proteins, with the same precision as previously known for small organic compounds. This enables the total chemical synthesis of proteins or the introduction of highly specific perturbations beyond the naturally encoded structural and chemical repertoire. For example, amino acids with an unnatural side chain or backbone composition can be incorporated to effect one-atom mutations in the protein of interest or to introduce biophysical probes and post-translational modifications. The same chemical tools are explored to build molecules with tailored activity on the basis of protein structures.

The introduction of native chemical ligation (NCL) made synthetic proteins of considerable size accessible and has had a dramatic impact on the field of protein chemistry. To quote Stephen Kent (Chicago), who introduced this technology:[1] "Once you have synthetic access to a protein, you are only limited by your imagination". This was nicely illustrated by Kent himself with a recent example from his group,[2] the 76 amino acid (aa) ubiquitin, which was assembled from three peptide fragments by NCL. Having synthetic access to the protein allowed the substitution of Gly35, which is found in the unusual left-handed conformation that, among the natural L-amino acids, only glycine can adopt, with an unnatural D-amino acid (D-Gln). A high-resolution crystal structure of [D-Gln]-ubiquitin revealed the predicted left-handed conformation of the D-isomer without further structural perturbations (Figure 1); this suggests that nature has indeed selected glycine at this position for its ability to adopt this conformation.[2]

One area in which (semi-)synthetic protein chemistry has proven to be especially valuable is in the field of post-translational modifications. Prenylation is essential for the proper function of many proteins, for example for the Rab proteins that are involved in vesicle transport. However, homogeneous samples of mono- and double-prenylated Rab cannot be obtained in significant yields by overexpression. Roger Goody (MPI Dortmund) showed how the NCL-related technique expressed protein ligation (EPL)[3] was used to join a recombinant version of Rab with a synthetic peptide that corresponds to the C-terminal sequence of the protein, which includes the prenylated side chains. Crystal structures of the prenylated Rab proteins complexed with RabGDI (GDP-dissociation inhibitor) provided insights into the hydrophobic isoprenoid-binding pocket of RabGDI and suggested a model for Rab extraction from the membrane by RabGDI.[4] That EPL has become an established technique was evident from the number of lectures in which it was discussed. Christian Becker (MPI Dortmund) has used both EPL and protein trans-splicing mediated by split inteins for the synthesis of lipid-modified prion proteins. Carlo Unverzagt (Bayreuth) described his progress towards the semi-synthesis of glycosylated RNase B, and a new split intein, which can be used to ligate synthetic probes to recombinant proteins, was presented by Henning Mootz (Marburg).

The formation of 4-hydroxyproline is an important post-translational modifica-

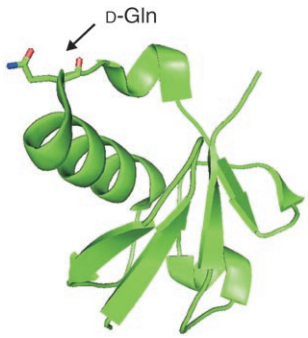

Figure 1. Crystal structure of [D-Gln35]-ubiquitin. The figure was created from PDB accession code 1YJ1 (see ref. [2])

[a] *Dr. G. Roelfes*
Department of Organic Chemistry
Stratingh Institute
University of Groningen
Nijenborgh 4, 9747 AG Groningen
(The Netherlands)
Fax: (+31) 50-363-4296
E-mail: J.G.Roelfes@rug.nl

[b] *Dr. H. D. Mootz*
Fachbereich Chemie/Biochemie
Philipps-Universität Marburg
Hans-Meerwein-Strasse, 35032 Marburg
(Germany)
Fax: (+49) 6421-2822191
E-mail: Mootz@chemie.uni-marburg.de

tion that contributes to the stability of the collagen triple helix. By chemical synthesis of the isoelectric 4-fluoroproline and its incorporation into synthetic collagen, a hyperstable triple helix was obtained.[5,6] This earlier report by Ronald Raines (Wisconsin–Madison) showed that inductive effects and not hydrogen bonds of the 4-hydroxyproline are important for stability. More recent results that were presented further deciphered the electronic influence of the stereogenic center at the 4-position and suggested how these findings could be exploited for a controlled self-assembly of collagen from different strands.[7]

Proteins can be turned into sensors by covalently attaching a fluorophore close to the substrate-binding site. In this fashion, Martin Webb (Imperial College, London) created phosphate sensors by linking the environmentally sensitive fluorophore coumarin to phosphate-binding proteins.[8] Sensing the production of orthophosphate allowed the study of ATP-dependent DNA helicases. In another example, a sensor for the ADP/ATP ratio was prepared by attaching a coumarin to a nucleoside diphosphate kinase.

A very powerful alternative to chemistry for the introduction of new functions in proteins is directed laboratory protein evolution. Dan Tawfik (Weizman) addressed the question of the time point of implementation of new functions in proteins during evolution. Selected mutations that increased the activity of PON1 hydrolase and other enzymes towards a promiscuous substrate only marginally affected the activity of the mutant proteins for its wild-type substrate. These results support the hypothesis that protein evolution proceeds through nonspecific intermediates, followed by gene duplication to diverge into robust new functions with subsequently acquired mutations.[9]

Catalytic antibodies have long been considered promising as useful tailor-made biocatalysts that can be raised for a wide range of chemical reactions. However, their lower catalytic activity compared to natural enzymes is an inherent problem and is, among other reasons, a result of the fact that they are selected for binding to transition state analogues and not for catalysis. Don Hilvert (ETH Zürich) laid out the structural basis for the Diels–Alderase activity of a catalytic antibody and showed how, by a combination of site-directed mutagenesis and directed evolution, the substrate binding and catalytic properties could be optimized.[10,11] More on the structure of evolved proteins was presented by Sheref Mansy (from the group of J. Szostak, Harvard Medical School), who discussed the solution structure of an artificial ATP-binding protein discovered by using in vitro ribosome display technology. In this same session, Robin Leatherbarrow (Imperial College, London) presented the development of an inhibitor for the 3C protease from the foot and mouth disease virus. Victor Bolanos-Garcia (University of Cambridge) shared new physical insights into the structural properties that are important for the function of BUB1/BUBR1 proteins.

Molecular recognition is of central importance in biological systems. Ernest Giralt (Barcelona) showed how principles of protein–protein interactions can be investigated with synthetic molecules. Oligomers of chiral, bicyclic guanidines[12] interact with peptides and proteins through shape and charge complementarity. One such oligomer recognized an anionic helical patch on the surface of the tetramerization domain of the tumor-suppressor protein p53.[13] Similar tetraguanidinium compounds also mediate internalization into cells and may serve as nonhydrolyzable drug carriers.[14] Synthetic mimics of discontinuous binding sites in proteins have been developed by Jutta Eichler (GBF Braunschweig) on cyclic peptide scaffolds and explored as vaccine candidates. A very different scaffold for the construction of protein mimics is made by gold nanoparticles that Raphaël Lévy (Liverpool) covered with peptides with the aim of introducing biomimetic functions.

Protein–carbohydrate interactions represent an exquisite example of molecular recognition between biomolecules. A main difficulty in the field of carbohydrates stems from their complex and heterogeneous structures. As Peter Seeberger (ETH Zürich) discussed in his lecture, the improved and automated synthesis of oligosaccharides by solid-phase synthesis[15] gives rise to fast access of large amounts of homogeneous samples. With these synthetic compounds at hand, biochemical studies can be performed, for example on the glycan-dependent gp120 protein interactions during HIV entry.[16] Importantly, they also provide a rational approach to the development of vaccines, as exemplified in a malaria model with the use of synthetic glycosylphosphatidylinositol, the toxine of the parasite,[17] and the synthesis of a tetrasaccharide antigen of *Bacillus anthracis*.[18]

When investigating the proteomic content in a biological sample, such as a certain tissue, it is crucial to distinguish between fractions of active and inactive enzymes. Nicolas Winssinger (Strasbourg) used a library of small peptide substrates for proteases that are linked to a fluorophore and a PNA tag. The latter served as a bar-code for the individual sequence of the peptide. Upon incubation with a mixture of proteases in the sample, the cleavage of peptides unmasks the fluorophore. The activity profile can then be spatially deconvoluted by reading out the fluorescence after hybridization of the PNA tags to a DNA microarray.[19] The high sensitivity is a great advantage of this method, but overlapping activities in the sample could be problematic. An alternative approach, presented by Benjamin Cravatt (Scripps Research Institute), overcomes the latter limitation and can also be performed in whole cells and even organisms. Here, the enzyme's active site is irreversibly labeled with a fluorophore, the protein mixture is separated on an SDS-PAGE gel, and the marked protein is identified by MS analysis.[20] The activity profiles from various human cancer cell lines can be clustered into phenotypically relevant groups to identify interesting proteins for further analysis.[21]

Kai Johnsson (EPFL Lausanne) showed how chemistry in living cells can be used to study proteins in their "natural" environment. Selective chemical labeling takes place in the cell on the small proteins O^6-alkylguanine DNA alkyl transferase (AGT)[22] and acyl-carrier protein (ACP), which are expressed in a fusion with the protein of interest. Various applications are possible, including the

study of protein–protein interactions and the creation of protein microarrays. Fluorescent labeling can be extended to time-dependent or orthogonal multicolor imaging, as demonstrated by visualization of the localization of cell-wall growth in budding yeast.[23] Cells can also be "miniaturized", an approach followed by Horst Vogel (EPFL Lausanne). He described the use of so-called native vesicles, that is, miniaturized cells containing receptors that are still functional.[24] These native vesicles are used as single attoliter-sized containers, but also on microarrays to study electrophysiology on a chip. Cell-penetrating peptides (CPPs) are one possible and relatively nontoxic way to bring a synthetic and hydrophilic cargo inside a cell. Ülo Langel (Stockholm) discussed applications such as delivery of antisense RNA and plasmids into cells and introduced an algorithm to predict CPP sequences.[25] Tom Muir (Rockefeller University) discussed aspects of chemical communication between bacteria, known as quorum sensing. The bacterial warfare between staphylococci is mediated by small peptides known as autoinducing peptides (AIPs), which contain a unique thiolactone structure. These peptides can act as agonists on their own receptor and as antagonist on other *Staphylococcus* strains and thereby switch virulence on or off.[26] Chemical synthesis of AIP variants allowed their structure–function relationships and the importance of hydrophobic interactions for binding to the receptor to be deciphered.[27]

Chemistry also provides many potential routes to external control of peptide and protein secondary structure, for example to modulate biological systems in a defined manner. Luis Moroder (MPI Martinsried) made the case for light as the ideal external trigger. Azobenzene is a well-known photochromic moiety that can be efficiently switched between the *cis* and *trans* isomers by using light of different wavelengths.[28] Incorporation of the azobenzene moiety into a β-hairpin peptide allowed photocontrol over the conformation of this secondary-structure element and might be the prelude to regulating the function of entire proteins in this manner.[29] Kerstin Rießelmann (from the group of K. Rück-Braun, TU Berlin) described a new class of photoswitchable amino acids based on hemithioindigo moieties, which could be used as a photoswitchable element in cyclic Grb2-SH2 antagonists. Natural enzymes that undergo photoswitching are, of course, perfectly adapted to the light source. Martin Engelhard (MPI Dortmund) reported on the conformational changes within the rhodopsin family of proteins in response to the signal. These processes were studied by using site-selectively introduced spin labels. The line broadening of the EPR signals indicated the mobility of the protein in this particular region.[30] Mudi Sheves (Weizman) presented the hypothesis that a large induced dipole is created in the retinal upon light absorption, which causes structural rearrangement. To illustrate this, locked chromophores were synthesized that cannot isomerize, but still cause movement within the protein.[31] Light is not only used for photoswitching, but is also an important source of energy. Light-harvesting proteins are crucial to collect this energy. Huub de Groot (Leiden) used solid-state NMR to study the direct contacts between light-harvesting complexes and the protein matrix by employing ^{13}C-labeled chlorophyll.[32]

A beautiful account of the application of physical organic chemistry to biological questions regarding a neurotransmitter's binding to its receptor was given by Dennis Dougherty (Caltech). Cation–π interactions are important in the binding of acetylcholine to the nicotinic acetylcholine (nAch) receptor.[33] Support for this was obtained by replacement of the key tryptophan residue involved in binding by fluorinated analogues, by using the in vivo nonsense-suppression methodology. In contrast, nicotine was found to bind in a different fashion, involving hydrogen bonding to the peptide backbone. Introduction of α-hydroxy amino acids, which give rise to ester linkages in the backbone, led to reduced hydrogen-bonding capability and, as a result, reduced affinity for nicotine. A long-standing problem is how ligand binding in the outer domain is translated into opening and closing of the channel inside the membrane. An essential proline residue was identified in the loop between the membrane helices 2 and 3 of the 5-HT3 receptor that is in contact with the ligand-binding domain. Exchange of this amino acid with various proline analogues suggested that prolyl *cis–trans* isomerization mediates the signal to open the pore of the channel (Figure 2).[34] Hagan Bayley (Oxford) showed how an α-hemolysin pore is exploited as an artificial nanoreactor, in which chemistry can be performed at the single molecule level. When arsenic compounds bind to an engineered Cys, individual bond-forming and -breaking reactions could be observed. A new development that was shown was the stepwise growth of a single polymer chain inside this pore.[35]

Sometimes chaperones are needed to achieve correct folding of a polypeptide chain into a functional protein. Cordelia Schiene-Fischer (MPI Halle) discussed the DnaK protein, which is a member of the Hsp70 class of chaperones. It was demonstrated that DnaK facilitates protein folding by catalyzing the *cis–trans* isomerization of non-prolyl peptide bonds, which is normally slow on the biological timescale.[36] Then the focus shifted to the early events in protein folding: first contact and nucleation. Thomas Kiefhaber (Biozentrum Basel) used unnatural amino acids based on xanthone and naphthalene to study these phenomena. Upon excitation, rapid triplet–triplet energy transfer can take place between the chromophores, provided there is van der Waals contact.[37] This could be used to determine the "speed limit" for protein folding, and it was found that the timescale of these initial events is in the nanosecond range. This methodology was applied to model peptides and the protein parvalbumin.

Backbone hydrogen bonding is an important aspect in protein stability, as is illustrated by the fact that, on average, in a protein, 1.1 such H-bonds exist per residue. To study them, Jeff Kelly (Scripps Research Institute) replaced the amide bond either with an ester group or with an *E* alkene, which is isosteric with a peptide bond. The power of this chemical approach was demonstrated with a study of the WW domain from the PIN1 protein, a 34-residue peptide that can be independently folded.[38] Replacement of

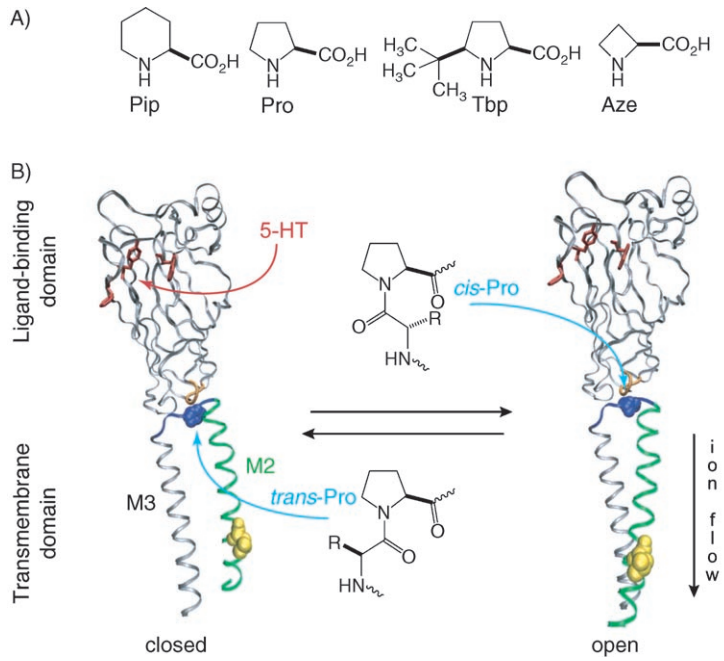

Figure 2. A) Proline analogues incorporated into the 5-HT_3 receptor. B) Model for the transduction of the ligand-binding signal to the gating of the pore by prolyl *cis–trans* isomerization. Illustration adapted from ref. [34].

peptide bonds involved in backbone hydrogen bonds crucial for stability with an ester or *E* alkene analogue resulted in unfolded protein. Quantification of the influence of amide-to-alkene mutation on the energetics of the transition state compared to the native state by ϕ_M analysis suggested that loop formation is an important first step in folding of this peptide. Finally, this methodology has been applied to gain more insight into the structural requirements of amyloid formation. By incorporation of an *E* olefin, in the form of a Phe–Phe isostere in an amyloid protein, the formation of spherical aggregates instead of the typical fibrillar structure was observed;[39] this demonstrates the importance of backbone H-bonding in amyloid structures.

In summary, chemical approaches can be extremely powerful for probing the molecular basis of protein function. Recent technological advances have largely expanded the protein space accessible for the synthetic chemist and further useful tools are likely to emerge. Most important for the successful marriage of chemistry and biology, however, will be the creativity and imagination to devise the right experiment. It might have been for this reason, that the conference excursion took the participants to the museum that is dedicated to, arguably, the most creative and imaginative artist this region has produced: Salvador Dalí. But also an exciting scientific meeting like this one, with stimulating talks and vivid discussions, certainly helps and should be followed up in the future.

[1] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776.
[2] D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakoff, S. B. Kent, *Angew. Chem.* **2005**, *117*, 3920; *Angew. Chem. Int. Ed.* **2005**, *44*, 3852.
[3] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.
[4] A. Rak, O. Pylypenko, T. Durek, A. Watzke, S. Kushnir, L. Brunsveld, H. Waldmann, R. S. Goody, K. Alexandrov, *Science* **2003**, *302*, 646.
[5] S. K. Holmgren, K. M. Taylor, L. E. Bretscher, R. T. Raines, *Nature* **1998**, *392*, 666.
[6] S. K. Holmgren, L. E. Bretscher, K. M. Taylor, R. T. Raines, *Chem. Biol.* **1999**, *6*, 63.
[7] J. A. Hodges, R. T. Raines, *J. Am. Chem. Soc.* **2005**, *127*, 15923.
[8] M. Brune, J. L. Hunter, J. E. Corrie, M. R. Webb, *Biochemistry* **1994**, *33*, 8262.
[9] A. Aharoni, L. Gaidukov, O. Khersonsky, S. McQ. Gould, C. Roodveldt, D. S. Tawfik, *Nat. Genet.* **2005**, *37*, 73.
[10] A. Piatesi, D. Hilvert, *ChemBioChem* **2004**, *5*, 460.
[11] J. Chen, Q. Deng, R. Wang, K. Houk, D. Hilvert, *ChemBioChem* **2000**, *1*, 255.
[12] B. P. Orner, X. Salvatella, J. Sanchez Quesada, J. De Mendoza, E. Giralt, A. D. Hamilton, *Angew. Chem.* **2002**, *114*, 125; *Angew. Chem. Int. Ed.* **2002**, *41*, 117.
[13] X. Salvatella, M. Martinell, M. Gairí, M. G. Mateu, M. Feliz, A. D. Hamilton, J. De Mendoza, E. Giralt, *Angew. Chem.* **2004**, *116*, 198; *Angew. Chem. Int. Ed.* **2004**, *43*, 196.
[14] J. Fernandez-Carneado, M. Van Gool, V. Martos, S. Castel, P. Prados, J. de Mendoza, E. Giralt, *J. Am. Chem. Soc.* **2005**, *127*, 869.
[15] O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **2001**, *291*, 1523.
[16] E. W. Adams, D. M. Ratner, H. R. Bokesch, J. B. McMahon, B. R. O'Keefe, P. H. Seeberger, *Chem. Biol.* **2004**, *11*, 875.
[17] L. Schofield, M. C. Hewitt, K. Evans, M. A. Siomos, P. H. Seeberger, *Nature* **2002**, *418*, 785.
[18] D. B. Werz, P. H. Seeberger, *Angew. Chem.* **2005**, *117*, 6474; *Angew. Chem. Int. Ed.* **2005**, *44*, 6315.
[19] N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick, J. L. Harris, *Chem. Biol.* **2004**, *11*, 1351.
[20] A. E. Speers, G. C. Adam, B. F. Cravatt, *J. Am. Chem. Soc.* **2003**, *125*, 4686.
[21] N. Jessani, S. Niessen, B. Q. Wei, M. Nicolau, M. Humphrey, Y. Ji, W. Han, D. Y. Noh, J. R. Yates III, S. S. Jeffrey, B. F. Cravatt, *Nat. Methods* **2005**, *2*, 691.
[22] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86.
[23] L. Vivero-Pol, N. George, H. Krumm, K. Johnsson, N. Johnsson, *J. Am. Chem. Soc.* **2005**, *127*, 12770.
[24] H. Pick, E. L. Schmid, A. P. Tairi, E. Ilegems, R. Hovius, H. Vogel, *J. Am. Chem. Soc.* **2005**, *127*, 2908.
[25] S. El-Andaloussi, T. Holm, U. Langel, *Curr. Pharm. Des.* **2005**, *11*, 3597.
[26] P. Mayville, G. Ji, R. Beavis, H. Yang, M. Goger, R. P. Novick, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1218.
[27] J. S. Wright III, G. J. Lyon, E. A. George, T. W. Muir, R. P. Novick, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16168.
[28] R. Behrendt, C. Renner, M. Schenk, F. Wang, J. Wachtveitl, D. Oesterhelt, L. Moroder, *Angew. Chem.* **1999**, *111*, 2941; *Angew. Chem. Int. Ed.* **1999**, *38*, 2771.
[29] S. L. Dong, M. Loweneck, T. E. Schrader, W. J. Schreier, W. Zinth, L. Moroder, C. Renner, *Chemistry* **2005**.
[30] E. Bordignon, J. P. Klare, M. Doebber, A. A. Wegener, S. Martell, M. Engelhard, H. J. Steinhoff, *J. Biol. Chem.* **2005**, *280*, 38767.
[31] G. Fan, F. Siebert, M. Sheves, R. Vogel, *J. Biol. Chem.* **2002**, *277*, 40229.
[32] A. J. van Gammeren, F. Buda, F. B. Hulsbergen, S. Kiihne, J. G. Hollander, T. A. Egorova-Zachernyuk, N. J. Fraser, R. J. Cogdell, H. J. de Groot, *J. Am. Chem. Soc.* **2005**, *127*, 3213.
[33] A. L. Cashin, E. J. Petersson, H. A. Lester, D. A. Dougherty, *J. Am. Chem. Soc.* **2005**, *127*, 350.

[34] S. C. Lummis, D. L. Beene, L. W. Lee, H. A. Lester, R. W. Broadhurst, D. A. Dougherty, *Nature* **2005**, *438*, 248.
[35] S. H. Shin, H. Bayley, *J. Am. Chem. Soc.* **2005**, *127*, 10462.
[36] C. Schiene-Fischer, J. Habazettl, F. X. Schmid, G. Fischer, *Nat. Struct. Biol.* **2002**, *9*, 419.
[37] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, T. Kiefhaber, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9597.
[38] S. Deechongkit, H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, J. W. Kelly, *Nature* **2004**, *430*, 101.
[39] Y. Fu, J. Bieschke, J. W. Kelly, *J. Am. Chem. Soc.* **2005**, *127*, 15366.

Received: December 16, 2005
Published online on February 2, 2006

